BBC
Research Department Report

TELEVISION MOTION MEASUREMENT FOR DATV AND OTHER APPLICATIONS

G.A. Thomas, B.A.(Hons), A.M.I.E.E.

Research Department, Engineering Division
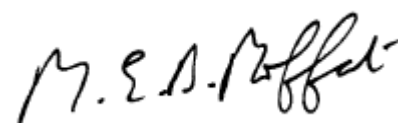**THE BRITISH BROADCASTING CORPORATION**

TELEVISION MOTION MEASUREMENT FOR DATV AND OTHER APPLICATIONS

G. A. Thomas, B.A.(Hons), A.M.I.E.E.

Summary

Almost all picture processing operations that work in the temporal domain can benefit from knowledge of the speed and direction of moving objects in the scene. Generally, motion vector measurement techniques that have been developed so far do not work well enough for broadcast quality applications. The Report reviews four existing techniques, and suggests some novel extensions to a technique based on phase correlation. The results of simulating this technique on a computer image processing system are reported Several specific applications have been investigated, including temporal standards conversion and bandwidth reduction in a system using DATV (Digitally Assisted Television). The results are encouraging, and suggest that the proposed vector measurement technique can be used successfully in critical broadcast quality applications.

Issued under the Authority of

Research Department, Engineering Division
BRITISH BROADCASTING CORPORATION

Head of Research Department

TELEVISION MOTION MEASUREMENT FOR DATV AND OTHER APPLICATIONS

G. A. Thomas, B.A.(Hons), A.M.I.E.E.

TELEVISION MOTION MEASUREMENT FOR DATV AND OTHER APPLICATIONS

G. A. Thomas, B.A.(Hons), A.M.I.E.E.

1. INTRODUCTION

There are many picture processing applications in which knowledge of the speed and direction of movement of all parts of the TV picture would be very useful. These applications include standards conversion, noise reduction, bandwidth reduction[1,2], and others where any sort of temporal interpolation is required. Even information concerning simple camera panning movements and the movement of larger objects in the scene would sometimes be useful.

Motion vector information is needed in these applications because the television signal is usually not filtered in the manner required by the Nyquist criterion prior to sampling in the temporal domain. Thus a moving television picture contains information aliased temporally. Far from being a disadvantage, the aliased nature of the signal in the temporal domain is an important feature in many cases, since it allows us to maintain a high degree of spatial resolution on moving objects. However, it also means that conventional linear interpolation techniques cannot be applied successfully in the temporal domain.

Many papers have been written on various motion measurement (or 'movement following') techniques. Most of the techniques involve dividing the picture into small blocks and calculating a motion vector for each block. Few of the papers give any clear indication as to how well the techniques work, as performance is often assessed in terms of a bandwidth reduction system into which the technique is incorporated. The results given suggest that most techniques do not work well enough for broadcast use, particularly for large or rapidly changing movements, or for scenes containing many objects moving separately. Ideally, a technique should be able to measure movements up to about 15 pixels per field period (about one second per picture width) for a standard TV signal, to an accuracy better than one pixel. It would also be useful to have vectors assigned to individual pixels rather than to blocks.

The Report gives a brief summary of motion measurement methods reported in the literature and suggests some novel modifications to one basic technique. The results of simulating this technique in software are reported, and some possible applications are discussed. The Report also describes how the technique can be used to provide motion vector information for a bandwidth reduction system based on the concept of Digitally Assisted Television[1,2,3].

2. A REVIEW OF PUBLISHED MOTION MEASUREMENT TECHNIQUES

Most motion measurement techniques discussed in the literature fall into four categories, namely methods based on spatio-temporal differentials, matching techniques, Fourier techniques and techniques based on feature extraction. The first two are usually applied on a block-by-block basis, blocks typically being about 16 pixels by 8 lines with an interlaced system.

These four types of technique are described in more detail below.

2.1 Techniques based on spatio-temporal differentials

These techniques are based on the assumption that the intensity variation across a TV field is a linear function of displacement. This is equivalent to assuming that the displacement to be measured is small compared to the wavelength of the highest image frequency component present. It is also necessary to assume that the brightness of objects does not change as they move. There are many examples in the literature of such techniques, for example Refs. 4 and 5.

The luminance difference between corresponding pixels in successive frames is calculated and summed over a block. The difference between adjacent pixels is also summed, in both the horizontal and vertical directions. The ratio between the frame difference and the horizontal and vertical element differences gives the horizontal and vertical shifts respectively, in units of pixels per frame.

Although such techniques work well for sub-pixel shifts, they fail for larger movements. It is possible to apply such methods recursively, by displacing the latest input picture by an amount corresponding to the estimated shift based on previous measurements. This can help the measurement converge to the correct value, although convergence can be slow and in some cases does not occur at all. Some recursive techniques update the displacement estimate on a pixel-by-pixel basis, and some reset this 'running estimate' at what is thought to be an object edge.

Even when these refinements are incorporated, spatio-temporal differential techniques tend not to work particularly well. They have the advantage of

being relatively simple to implement, although once a number of recursive refinements have been included, the complexity can increase substantially. They are prone to failing in areas containing significant movement[6] and this makes them unsuitable for many applications, since it is precisely in those areas that accurate motion measurement is most important.

2.2 Techniques based on matching

This class of technique works by dividing the picture into small blocks and summing the mean square difference (or similar function) between each pixel of corresponding blocks in adjacent fields. This calculation is performed with several different spatial offsets between the blocks and the offset that gives the minimum error is taken as the motion vector for that block. The way in which trial offsets are chosen varies from method to method.

In one implementation of this technique[7], 25 different offsets are chosen for each block and the one giving the least match error is chosen. Offsets are formed by adding the motion vector calculated for the corresponding block in the previous field to a fixed set of 25 vectors. This means that the method only has to detect changes in motion vectors. The set of 25 vectors is chosen to give reasonable resolution (one pixel) for small movement differences, while still covering movement differences up to 5 pixels per field period.

There is no clear indication of how well this technique works, since the only assessment made is of the bandwidth reduction system in which this technique was incorporated. From this it appears that the system fails for rapidly accelerating objects (a failure common to all systems that look for changes in motion vectors), and when parts of the picture within a block are moving in different ways (a failure common to all systems based on blocks). This method cannot measure sub-pixel movement and so could produce impairments with such motion.

Another implementation of this technique[8] uses a 'logarithmic' search procedure to find the displacement vector that gives the minimum square error between the displaced blocks. This requires the assumption that the error decreases monotonically as the displacement vector converges to the correct value. Although this reduces the number of comparisons required it is likely that the assumption will become invalid for large movements.

The effectiveness of this method is again judged only in terms of the performance of a bit rate reduction system in which the technique is incorporated. The test material used contained only simple movement. It is likely that the method will fail for large movements, and it will also suffer from the problem of parts of one block moving in different ways.

Although this class of technique can generate a large number of motion vectors it is not always an advantage. In one implementation[9], it was found to be useful to limit the vectors actually used to those that occurred most frequently in the picture. This reduced the number of incorrect vector assignments that occurred.

2.3 Fourier techniques

This class of technique has been used in the past for image registration problems[10]. In this application, the technique involves correlating two images by first performing a two-dimensional Fourier transform on each image, multiplying together corresponding frequency components, and performing a reverse Fourier transform on the resulting array. The result is an array of numbers (a 'correlation surface') which will have a peak at the coordinates corresponding to the shift between the two pictures.

Not only does the use of Fourier transforms reduce the amount of calculation required compared to performing a correlation in the spatial domain, but it also enables filtering to be performed on the correlation surface. In particular, the sharpness of the peak can be significantly increased by normalizing the amplitude of each frequency component prior to performing the reverse transform. If G_1 and G_2 are the discrete two-dimensional Fourier transforms of the two successive images, then the complex array Z is calculated at every spatial frequency (m,n) thus:

$$Z(m,n) = \frac{G_1(m,n)G_2{}^*(m,n)}{|G_1(m,n)G_2{}^*(m,n)|}$$

The correlation surface is given by the inverse Fourier transform of Z, which will only have real components. Such a correlation process is known as a 'phase correlation', since the normalizing process results in only the phase information being used. As all positional information is contained in the phase of the spatial frequencies making up an image, this technique isolates the required information, and is not confused by brightness changes in the scene. It also has a good noise immunity. In the case of global movement, the correlation surface consists of a sharp peak, or delta function, situated at coordinates corresponding to the displacement.

It has been reported[10] that this type of technique is capable of measuring very large shifts (many tens of pixels) to an accuracy better than a tenth of a pixel, by interpolating the correlation

surface. However, as it stands, the method is only capable of measuring global motion, and any slight rotation of the picture can reduce the height of the correlation peak significantly.

Although it is possible to obtain even greater accuracy by using other image registration algorithms[11], these are generally not as robust and do not have as good noise immunity as phase correlation. In any case, the measurement accuracy reported[10] is adequate for all the picture processing applications under consideration in this Report.

2.4 Techniques using feature extraction

This class of technique is often applied to problems such as the determination of the three-dimensional structure of a scene from a number of photographs taken from different locations[12]. The basis of these methods is to identify particular features in the scene (often edges or corners of objects), and follow the movement of these features from one picture to the next. This provides motion information at various points in the picture, and an interpolation process is used to assign motion vectors to the remaining picture areas.

One way of measuring the movement of the edge or corner features is by first applying a high pass filter to the image to isolate the edge, and then using techniques based on the spatio-temporal differential method described in Section 2.1 to measure the amount of movement. The edge information can be smoothed with a low pass filter to reduce the effects of noise and enable larger movements to be measured.

In another implementation of a feature extraction technique[13], features were extracted manually from the scene (in this case they were particular blood vessel junctions on an X-ray picture of a heart). The movement of these features was tracked using an algorithm similar to the phase correlation method described above. This allowed the three-dimensional movement of a beating heart to be measured.

This class of technique is useful for specialized scene analysis tasks such as those described above, but is not often used to measure motion in more general scenes. As these techniques rely on the extraction of particular features from the scene (such as edges), they can fail to measure the correct velocity in picture areas that do not contain such features. They have been applied to more general scenes[14], but this failing is apparent in the results presented. An example of the type of picture material that would probably cause such techniques to fail is a moving area containing fine detail, such as a horizontal camera pan across grass.

2.5 Summary of published techniques

None of the techniques discussed above are 'ideal' for all applications requiring motion measurement. Spatio-temporal gradient methods do not perform well with movements much over one pixel per field period, although they are fairly easy to implement in hardware. Block matching algorithms generally perform better, although hardware implementation can be difficult (almost all the work described above was carried out using computer simulation). Fourier techniques appear to provide the most accurate measuring ability over a very wide range of motion magnitudes, although they generally require large blocks to work on. The fundamental drawback with any block-based system is that problems arise when parts of a block are moving differently. Feature extraction techniques tend to fail in picture areas devoid of recognisable edges.

In applications such as bandwidth reduction systems where the motion information is used as the basis of a predictive coder, some of the failings of the motion measurement techniques discussed above are not too critical. For example, if a block matching algorithm finds a displacement vector that gives a minimum error over a localized region, this may be sufficient as a predictor even if it does not correspond to the actual motion vector.

However, in applications involving temporal interpolation (such as standards conversion), it is important that the measured motion vector corresponds to the actual motion rather than just pointing to a similar looking area in the next picture. In such applications it is also desirable if vectors are assigned on a pixel-by-pixel basis, so that it is not necessary to assume that all picture material within a block is moving in the same way.

3. A SUGGESTION FOR AN IMPROVED VECTOR MEASUREMENT TECHNIQUE

An ideal vector measurement method would have the accuracy of the Fourier technique described above, coupled with the ability of block matching algorithms to measure the motion of many separate objects in a scene. Such a method should also be able to assign vectors to individual pixels if required. A possible way of realising a method that may be able to approach these ideals is as follows.

In the first stage of the proposed vector measurement process, the input picture would be divided into fairly large blocks, maybe 64 pixels square or even bigger. A phase correlation would be performed between corresponding blocks in successive pictures, resulting in a number of correlation surfaces

describing the movement present in different areas of the picture. Each correlation surface would be searched to locate not one, but *several* dominant peaks resulting from the motion of objects within each block. Thus, using this novel approach, several motion vectors could be measured by each correlation process. The result of this stage of the process would be a list of motion vectors likely to be present in the picture, on an area-by-area basis. The correlation surface could be interpolated to provide vectors of sub-pixel accuracy.

The second stage of the process would involve taking the list of possible vectors measured in the first stage, and assigning them to appropriate areas of the picture. This assignment would be done by shifting the input picture by each vector in turn relative to the previous picture, and calculating the match error at each pixel. There are several ways in which this error could be calculated; the simplest one would probably be to calculate the modulus of the luminance difference. This process would produce an 'error surface' for each trial vector that would indicate how well the vector fitted all parts of the picture. It may be advantageous to apply a spatial filter to the match error to reduce the effect of noise. The vector giving the smallest match error would be assigned to each area. Areas of the picture for which no vector gave a good match would probably correspond to erratic motion or uncovered background, and could be dealt with appropriately. For areas of erratic motion, this could involve reverting to a different motion measurement algorithm, trying out additional motion vectors, or interpolating vectors from surrounding areas of the picture. If the area was thought to correspond to uncovered background, more elaborate action might be required, depending on the use to which the motion vector information might be put. A possible way of dealing with uncovered or obscured background in the context of temporal interpolation is discussed in Section 4.2.3.

Thus the proposed technique is similar to the block matching algorithms discussed above, except that the number of trial displacements is limited to those measured in the phase correlation process. This allows the number of trial vectors to be kept to a minimum while still enabling large displacements to be measured accurately. Also, it is no longer necessary to assign vectors on a block-by-block basis; the area of the picture used to determine if a vector fits or not can be smaller because the number of trial vectors is limited, making it easier to distinguish between them.

This vector measurement technique could be tailored to a particular application by changing the size of the measuring blocks. If it was only necessary to measure the motion vectors of the major objects in the picture, only one measuring block might be used. Similarly, if it was only necessary to measure global movement, satisfactory results could probably be obtained by performing one-dimensional correlations both horizontally and vertically, having first summed picture elements in both directions.

The technique is likely to perform better with translational movement than it is with zoom and rotational movement. These types of movement produce a continuous range of velocity vectors, only one of which would be measured per measuring block. This may prove adequate if there are enough measuring blocks in the picture. However, the number of measuring blocks that can be used is limited by the minimum size of each block. The dimensions of a measuring block must be at least twice the size of the largest movement expected in each dimension, in order that there is a large amount of overlap between picture material in corresponding blocks in successive pictures.

4. RESULTS OF A COMPUTER SIMULATION OF THE PROPOSED MOTION MEASUREMENT ALGORITHM

In order to evaluate the algorithm described above a series of simulations were carried out. The aim of these simulations was to find out how well the algorithm could be made to work, without necessarily being restricted by techniques that would be easy to implement in hardware. At a later stage of the work, it is planned to carry out further simulations in order to find out how much the 'ideal' algorithm can be simplified to make it suitable for a real-time hardware implementation for a given application.

The image processing system used for these simulations consisted of a VAX 11/750 computer coupled to purpose-built RAM based picture stores. These stores could hold the equivalent of 12 full-size monochrome pictures, but could be configured to display sequences of 48 quarter-size monochrome pictures, as well as shorter sequences of colour pictures (held as either *RGB* or *YUV* components). Two short moving test sequences were available, stored digitally on Winchester discs connected to the VAX. There were also a number of still test pictures available.

All simulation work was carried out on short monochrome sequences, with pictures roughly one quarter the size of a full picture.

4.1 Investigation of vector measurement

The first part of the investigations examined the 'vector measurement' stage of the algorithm. The aim of this part was to investigate the accuracy of the phase correlation technique, and to see how it

depended on the size of the displacement, the number of moving objects, the amount of noise, and so on.

4.1.1 Measurement accuracy for simple panning movements

Initial investigations dealt with simple panning movements where the whole of the picture moves together. Two picture portions were correlated, both taken from the same larger picture but at slightly different locations. Both picture sections were 64 pixels square, luminance only, and were taken from the test picture 'Formal Pond' which contained a reasonable amount of detail. Fig. 1 shows this picture and indicates the section used.

picture portions. The location of the peak has moved, and the fitting of a quadratic curve independently in the x and y directions gave the peak location as (9.99,0.02). Hence the shift has been measured to an accuracy of a few hundredths of a pixel in ten pixels. The height of the peak has diminished and noise has been introduced; both effects can be attributed to the revealed and obscured material at the edges of the picture.

Similar experiments were tried with a large range of shift values. Fig. 4 shows how the height of the peak corresponding to the shift reduced in size as the amount of uncovered and obscured background increased. The peak could be detected for as little

Fig. 1 - The test picture 'Formal Pond' showing the 64 by 64 pixel portion used in simple panning movement measurement investigations.

Fig. 2 shows the correlation surface obtained when two identical picture portions were correlated. The surface has been interpolated to show points corresponding to half integral pixel shifts; the 'ringing' around the central peak shows the impulse response of the interpolator. This interpolation was performed by applying a window to the phase array and padding it with zeros prior to performing the inverse Fourier transform. Fig. 3 was produced in the same way but with a horizontal shift of ten pixels between the two

overlap as 10 pixels (15% of the block width). Fig. 5 shows how the accuracy of the panning movement measurement depended on the shift size, for both integral and half-integral pixel shifts. In contrast to the results of the previous paragraph, the surface was not interpolated to produce half-integer pixel shifts before peak detection. This causes a loss of accuracy for non-integer shift values. Even so, measurement accuracies of the order of 0.1 pixel or better were obtained for shifts up to about 45 pixels (30% overlap).
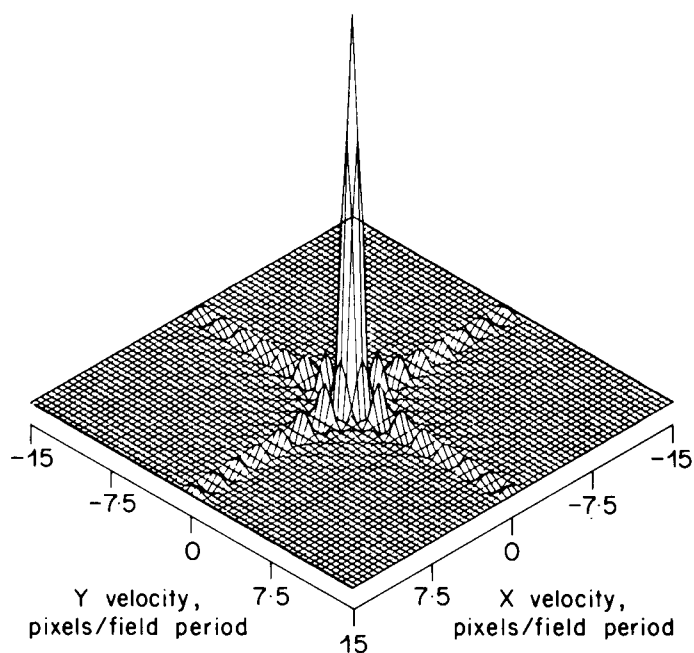
Fig. 2 - *Correlation surface for a stationary picture (points corresponding to half integral pixel shifts have been interpolated).*
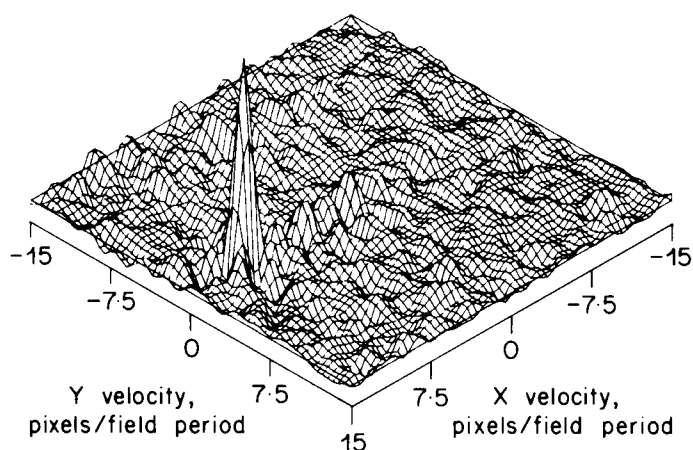

Fig. 3 - *Correlation surface for a simple panning movement with a horizontal shift of ten pixels between successive pictures.*

Fig. 6 shows the measurement accuracy for small shifts (up to about 5 pixels). The periodic nature of the error is again due to the performance of the interpolator; the improvement gained by interpolating half-integral points on the correlation surface prior to fitting the quadratic curve can clearly be seen. However, even this improved interpolation method gave errors of the order of 0.1 pixel for shifts of the order of 1 pixel or less. This error can be attributed to the lack of any windowing on the input picture, and is discussed in more detail in the next sub-section.

The experiments above were repeated with picture material containing less detail. A portion of the test picture 'Young Couple' shown in Fig. 7, was
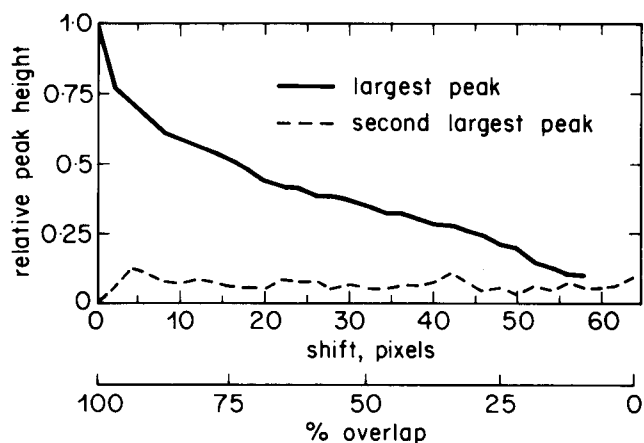

Fig. 4 - *Correlation peak height as a function of shift size for a picture portion 64 pixels square from 'Pond' (integral pixel shifts only).*
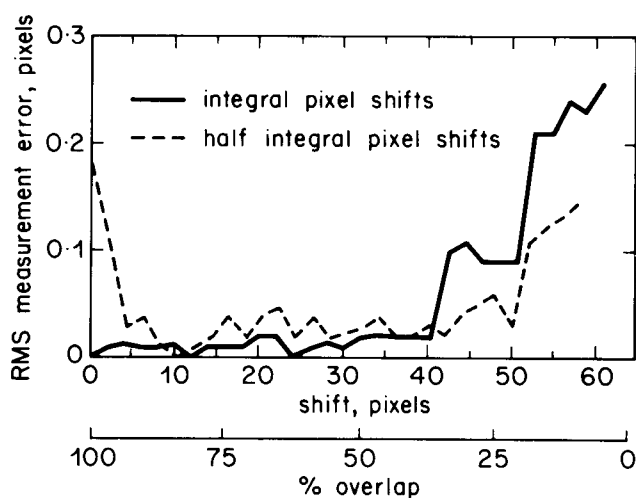

Fig. 5 - *Measurement error as a function of shift size for a picture portion 64 pixels square from 'Pond' (large movements).*

chosen. Fig. 8 shows the measurement error for large shifts, and Fig. 9 (a) shows the error for small shifts. The errors are generally slightly larger than those obtained when using the detailed picture portion (RMS measurement error of 0.120 pixel compared to 0.048 pixel for shifts up to 5 pixels using the better interpolation technique). However, the reduction in accuracy in such areas is not likely to present significant problems. For completeness, Fig. 9 (b) shows how the measurement accuracy was improved by windowing the input picture portion. In this case the use of windowing, to be discussed in more detail below, together with performing the interpolation by fitting a quadratic curve to interpolated half-integer pixel points gave an RMS measurement error of only 0.015 pixel.

These investigations showed that the technique was capable of accurately measuring shifts up to about 70% of the block size (about 45 pixels for blocks that were 64 pixels square). Shifts larger than half the
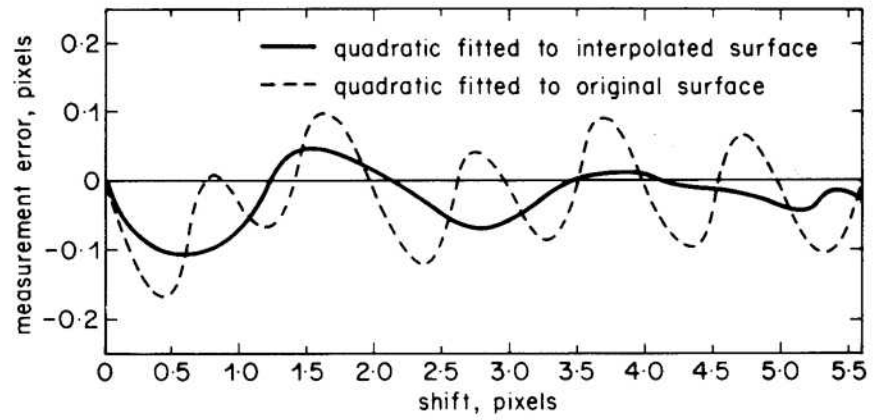
Fig. 6 - Measurement error as a function of shift size for a picture portion 64 pixels square from 'Pond' (small movements).

Fig. 7 - The test picture 'Young Couple' showing the portion used in panning movement measurement investigations.
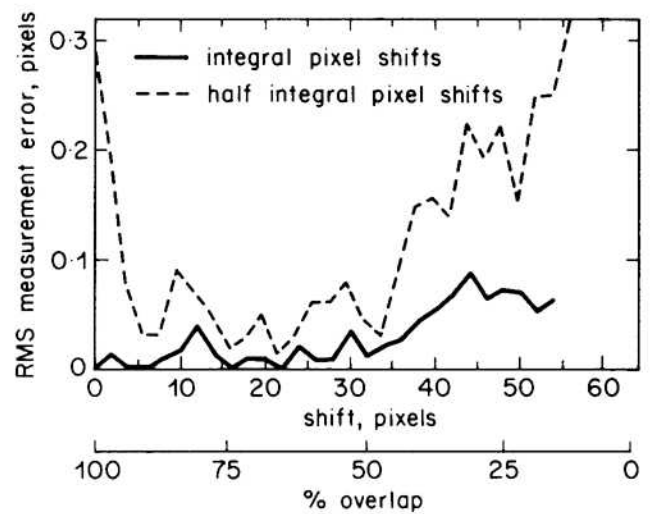

Fig. 8 - Measurement error as a function of shift size for a picture portion 64 pixels square from 'Young Couple' (large movements).

block size produced peaks at locations corresponding to smaller shifts in the opposite direction, since the axes of the correlation surface only ranged from −32 to +31 pixels horizontally and vertically. Thus a shift of +34 pixels appeared as a shift of −30. So in order to use blocks 64 pixels square to measure shifts greater than 32 pixels it would be necessary to have additional information to determine the direction of the shift. If it was necessary to measure such large shifts in a practical application, it may be possible to detect when a component of a motion vector of an object exceeded 32 pixels by following its motion as it accelerates up to and beyond that speed.

The method used above whereby the correlation surface is generated with interpolated points at half integral shift values is probably not the ideal method to use. The spurious peaks produced by the 'ringing' could mask peaks produced by small moving objects, and in any case it is unnecessary to interpolate points over the whole surface since it is only the area around a peak which is of interest. However, it is convenient to use this interpolation method for the purposes of displaying a correlation surface.

Similarly, it is useful to fit a quadratic curve to points on the correlation surface rather than a higher order polynomial because the location of the maximum can then be found uniquely and explicitly. A higher order polynomial could have a number of maxima, and solving for the location of these maxima becomes significantly more complex as the order increases. This is why it is better to perform the interpolation in two stages — first interpolating additional points using a conventional type of filter, then fitting a quadratic to the three interpolated points around the maximum.

4.1.2 The use of windowing on the input picture

It is possible to remove some of the noise introduced on the correlation surface by applying a 'windowing' function to the input picture portion. This has the effect of making the picture portion fade to mid-grey around the edges, so new picture material that appears at the edges contributes less to the correlation process. Thus the noise caused by the revealed and obscured background during camera pans is reduced.
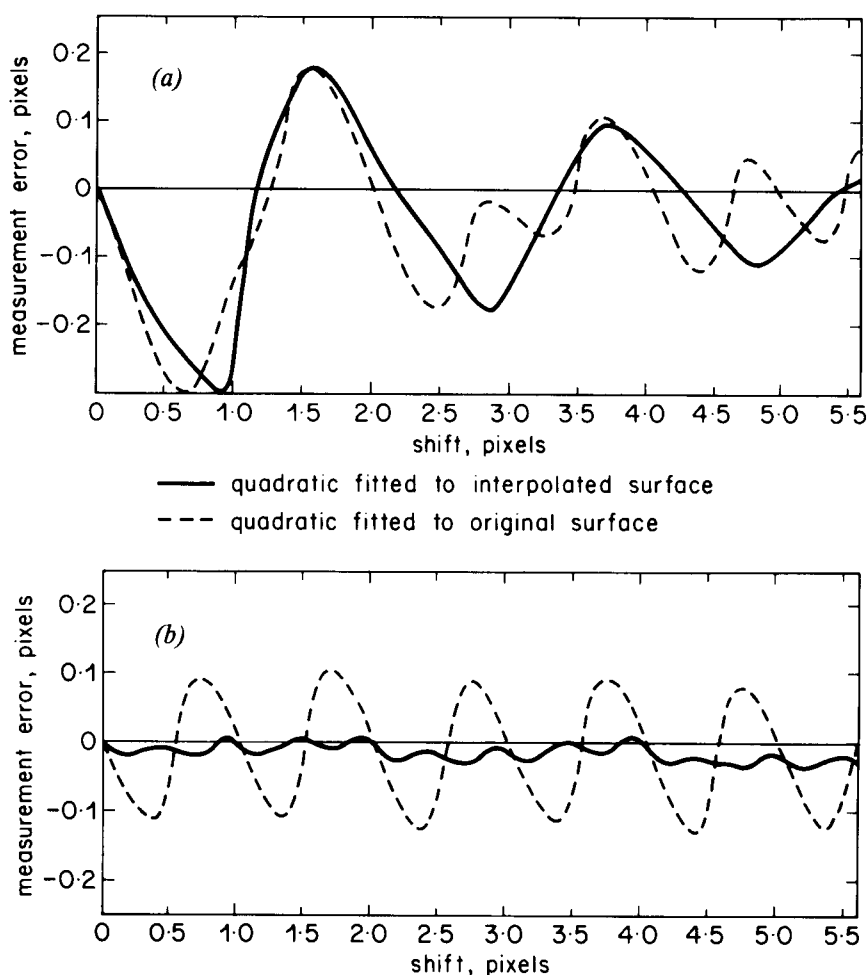

Fig. 9 - Measurement error as a function of shift size for a picture portion 64 pixels square from 'Young Couple' (small movements), (a) without and (b) with a windowing function applied to the picture portions.

Fig. 10 shows a correlation surface from the same panning movement that produced Fig. 3, but with the input picture portion windowed with a raised cosine window. This made the picture fade to grey (zero) around the edges; the overall dimensions of the windowed input picture portion plus its grey surround were 256 pixels square. A comparison of Figs. 3 and 10 shows that this windowing process has almost doubled the signal-to-noise ratio.

Another improvement gained by the use of windowing is the improvement of measurement accuracy for small movements, as mentioned in the previous sub-section. The reason for this is as follows. If the input picture portions are not windowed, there will be sharp luminance transitions at the edges of the block, as the left and right (and the top and bottom)

edges effectively join each other due to the periodic nature of the Fourier transform. The high-frequency components due to these transitions cause a spurious peak at zero displacement, as the transitions appear in the same places in both picture portions. This can have the effect of increasing the height of any peak at zero displacement.

For example, in the case of a shift of 0.5 pixel, the shift would be measured as being less than 0.5 pixel when no windowing is performed. This problem can be seen in Figs 6 and 9(a), where even the better interpolation technique underestimated such shifts by about 0.1 and 0.2 pixel respectively. Fig. 9(b) shows the improvement gained by windowing; the measurement error has been reduced from 0.21 pixel to 0.01 pixel. It should be noted that the severity of this problem depends on the picture material, particularly on the degree of difference between opposite edges of the picture portion.

The price to pay for the improvement gained by this windowing process is a doubling of the size of the transforms required in each direction. Also, since the input picture portion was windowed, it only effectively allowed velocities present in a smaller area of the picture to be measured. If this windowing technique were to be used, it would be necessary to have overlapping measurement blocks, which would increase the computational requirements by a further factor of roughly 4. For these reasons, the improvement gained is not likely to be warranted by the increased computation in many applications.

Simpler windowing techniques may be adequate for some applications. For example, the measurement accuracy could probably be improved by applying a spatial filter to the edges of the picture portion in such a way as to smooth out the effective joins between left and right hand edges (and top and bottom). This would remove any sharp transitions at the edges, and eliminate the spurious peak at zero velocity.

4.1.3 The effect of noise

The way in which noise on the input pictures affects the results of a phase correlation was investigated. Noisy pictures were generated by adding random amounts (with a given peak value) to the luminance level of each pixel. A correlation was performed between two pictures shifted by 10 pixels (as used to generate Fig. 3), with different levels of noise. Fig. 11 (a) shows how the level of noise affected the peak heights, and Fig. 11 (b) shows the effect on the accuracy. Noise levels below about −20 dB did not have a significant effect on the measurement process. This suggests that noisy input pictures do not present a significant problem to this
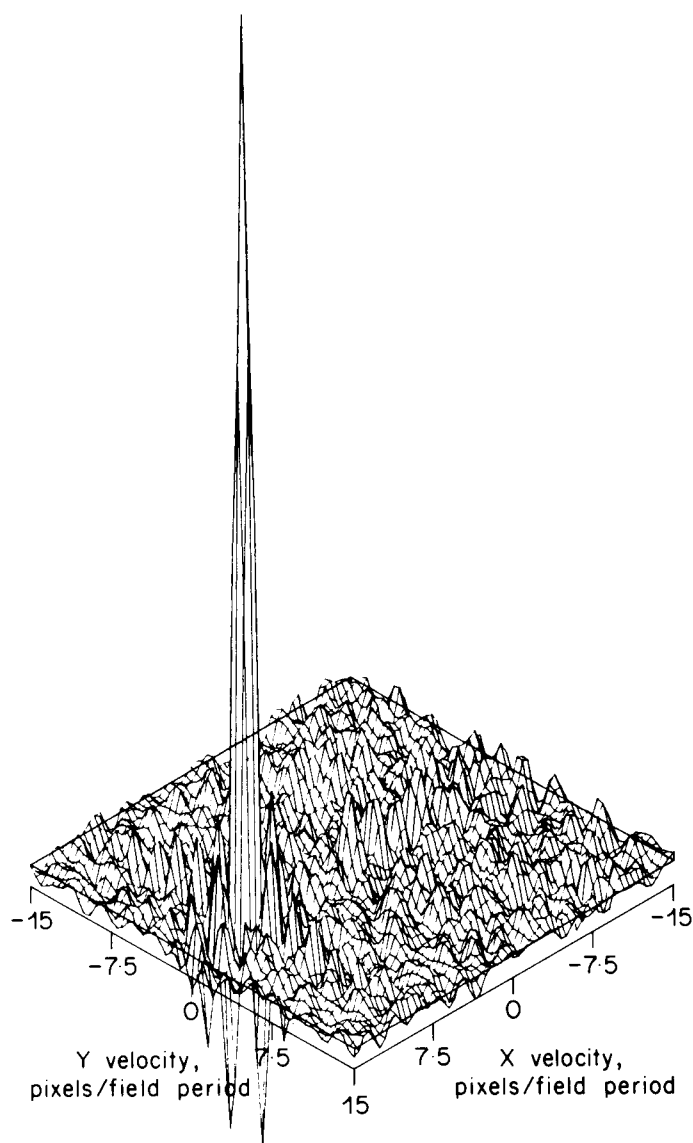
Fig. 10 - Correlation surface for a panning movement, with a horizontal shift of ten pixels between successive pictures. The picture portions were windowed with a raised cosine window and surrounded with mid-grey prior to correlation.

part of the motion measurement process. Indeed, the results imply that it may be sufficient to use the correlation process on picture data quantized to maybe only 4 bits in many cases, without seriously affecting the accuracy of the results.

Although noise on the input pictures may not present a problem, noise on the correlation surface can come from other sources, notably uncovered or obscured background. As will be seen in the following sub-section, this potentially can cause problems in measurement blocks containing several moving objects. Consequently, it is worth considering methods for reducing the level of noise.

It is possible that some improvement in signal-to-noise ratio may by obtained by applying various kinds of filtering to the correlation surface. Filtering in the frequency domain can be carried out by applying a weighting function to the phase array before performing the reverse transform. The weighting function could contain both a signal-independent part and a signal-dependent part.

A signal-independent filter which reduces the amplitudes of high frequency components can not only help to reduce noise, but can also provide the phase correlation algorithm with some immunity to geometrical distortions[10].

A signal-dependent filter could be used to reduce the amplitudes of frequency components that were of low amplitude in the input pictures. Such frequencies will contain little information and so will mainly contribute noise to the correlation surface. It would be desirable to limit the amplitudes of frequency components only if their amplitudes in the input picture were very low; other frequencies should all be given equal weighting. If each frequency component is given a weighting directly proportional to its amplitude in the input pictures, the correlation process becomes cross-correlation, and the sharpness of the peaks reduces significantly.

A temporal filter could also be applied to the correlation surface. This would reduce the effect of noise, while peaks that were nearly stationary would remain unchanged. This relies on the fact that few objects accelerate significantly between successive pictures. The response of the temporal filter would be chosen to provide adequate noise reduction without seriously impairing measurements of accelerating objects.

4.1.4 Scenes with more than one motion vector

Once the technique had been shown to work well for simple panning movements, pictures with an object moving over a background were investigated. The aim of these investigations was to see if the technique could accurately detect several moving objects by producing several peaks.

A portion of a picture 64 pixels square was extracted from the test picture 'Young Couple', and a 32 by 32 pixel portion from another picture (the blackboard cross from BBC 'Test Card F') was inserted centrally. The edges of the insert were 'blended' with the background over a distance of three pixels so as not to produce any artificially high frequencies in the resulting picture. A second similar picture was formed, but with the inserted portion shifted three pixels to the left and three down. Fig. 12 shows one of these composite pictures.

Fig. 13 shows the resulting correlation surface. The interpolated peak locations were accurate to 0.01 pixel. The relative heights of the peaks reflect the relative areas of moving object and background, although the ratio of heights is less than the ratio of object areas (1.6 compared to 4). This discrepancy is partly due to the fact that the height of a peak depends on the spatial frequency content of an object as well as its area. The noise around the peaks is due to the obscured and revealed background around the inserted picture portion.
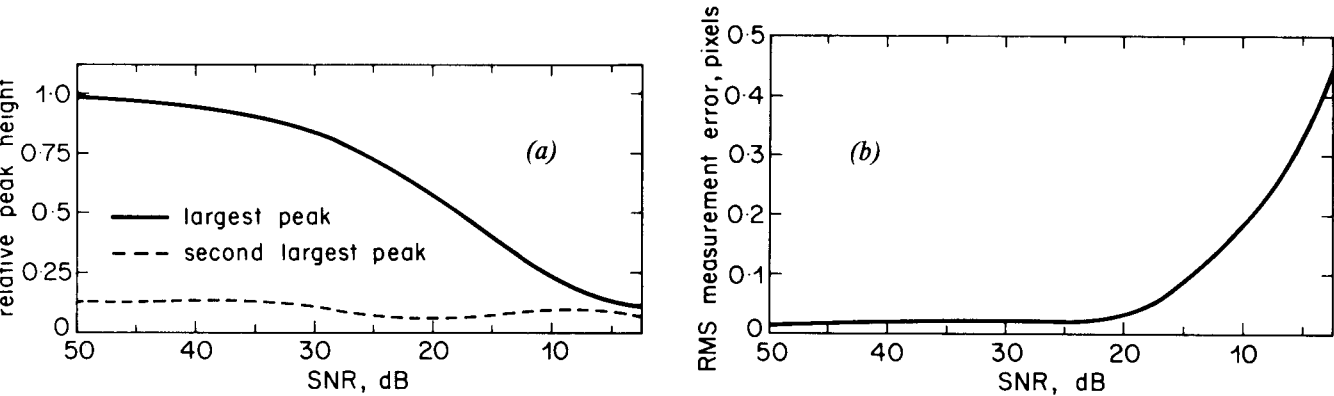

Fig. 11 - The effect of noise on (a) correlation peak height and (b) measurement error.
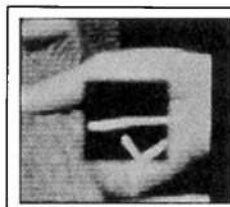
Fig. 12 - A composite picture generated from parts of BBC 'Test Card F' and 'Young Couple' used to investigate the measurement of two motion vectors in one correlation operation.

shows a vintage car which is moving towards the camera and slightly to the left. In the background there is a barred gate which is moving fairly rapidly to the right. The camera itself is performing a slow zoom out. A phase correlation was performed between a 64 by 64 pixel portion of two successive fields. The portion of the picture selected was the top left hand part of Fig. 14, showing the gate and the background. Fig. 15 shows the resulting correlation surface. The two large peaks correspond to velocities of (0.03, -0.42) and (2.38,-0.25). Measurements suggested that the speed of the gate was about 2.40 pixels per field period, which agrees very well with the value measured by phase correlation. Since the correlation was performed between two successive interlaced fields, a vertical shift of half a pixel would be expected, which agrees reasonably well with the vertical shifts measured. The heights of the two large peaks are roughly the same, reflecting the fact that the gate and background occupy areas of roughly the same size.
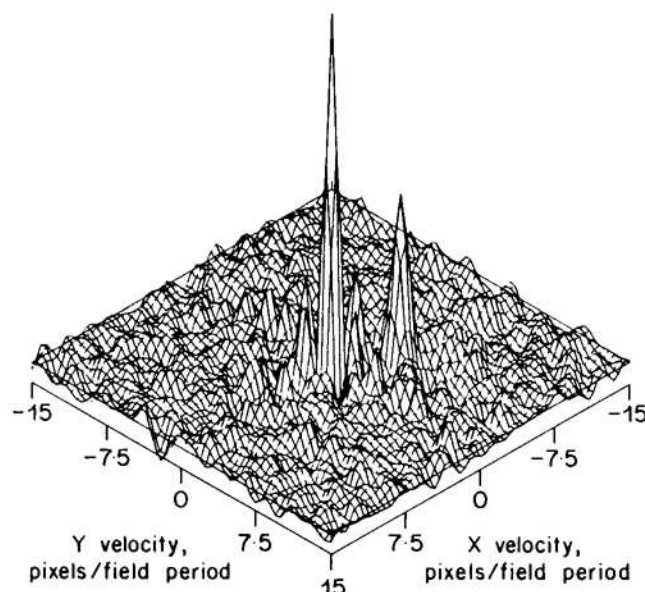

Fig. 13 - Correlation surface for an object moving over a stationary background.

As the phase correlation technique had no problems accurately measuring this motion, some more severe experiments were tried. The aim of these was to find out how small a moving object had to be before its motion vector could no longer be measured. A picture portion of various sizes was 'moved' with a shift vector of 5 pixels down and 5 to the right over a background picture, which was 64 pixels square. The motion vector of the object was accurately measured for objects as small as 2 pixels square, which was quite surprising, particularly as the object had moved more than twice its own length. Investigations with two objects moving over a stationary background showed that three peaks (one for the background and two for the objects) could usually be detected. In some cases, though, the peaks were becoming obscured by noise (probably due to uncovered background). These investigations did not use any of the techniques discussed earlier for improving the signal-to-noise ratio however, and it would probably be advantageous to apply some of these if it was necessary to detect many peaks in one measurement block. All investigations described above were carried out using computer generated moving pictures (constructed with portions of real pictures). The technique was also used to measure velocities present in a real sequence. Fig. 14 shows a picture from this sequence. The sequence

Fig. 14 - A picture from the 'Voiture'* test sequence showing the portion of the picture used to produce Fig. 15.
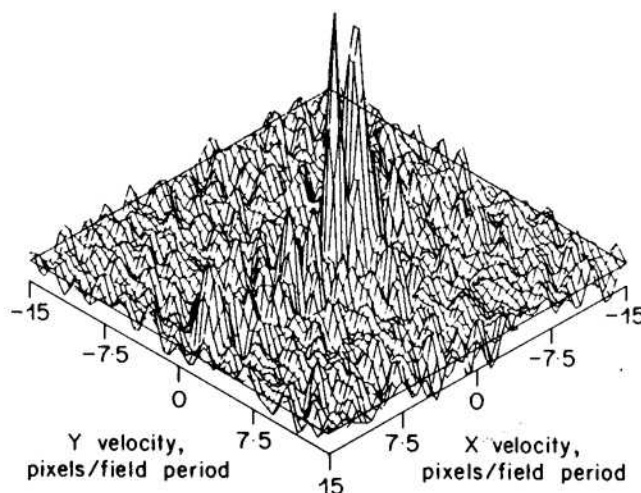

Fig. 15 - Correlation surface for a part of the 'Voiture' sequence for the small rectangular area indicated in Fig. 14.

* The Voiture sequence data was kindly provided by CCETT, Rennes, France. CCETT is the Joint Research Centre of TDF and the French PTT.

4.1.5 Summary of motion detection investigations

The experiments described above show that the phase correlation technique does indeed work as well as claimed[10]. For simple panning and one moving object, typical vector measurement accuracies of 0.05 pixel can be obtained by fitting a quadratic curve to the surface once the points corresponding to half integral shifts have been interpolated with a good interpolator. Higher accuracies (of the order of 0.015 pixel) can be obtained by windowing the input picture portions. This is an improvement over the accuracy reported in Refs. 10 and 11. An accuracy of about 0.2 pixel can be obtained by simply fitting a quadratic to the uninterpolated correlation surface. The accuracy of the technique was largely unaffected by noise, and it was possible to detect very small objects.

In most cases it was possible to detect three independently moving objects, although in some cases, peaks were obscured by noise. Several ways in which this noise may be reduced have been outlined.

Although all the investigations described above were carried out on blocks 64 pixels square, the technique can be applied to almost any block size as long as blocks are not so small that objects move completely out of a block between measurements. The block size used would depend on the application.

4.2 Investigation of vector assignment

The aim of this part of the investigation was to simulate the second stage of the motion measurement process described in Section 3, namely assigning the principal motion vectors to particular areas of the picture. The first problem that had to be dealt with was to find a way of detecting how well the vector measurement and assignment had been carried out. It was decided to do this by using the motion vectors to interpolate temporally a picture between the two input pictures. This was not only a stringent test of the method, but would also show what sort of results could be obtained if the technique was used in an application such as improving the motion portrayal of film, by generating intermediate pictures.

A computer program was developed that could generate the odd fields of a sequence by temporal interpolation between the even fields (or vice versa). The program allowed the user to change various parameters associated with the method, such as the size of the input picture, the size of the blocks on which the correlation was performed, the number of vectors extracted per block, and so on.

Ideally, an interlaced-to-sequential conversion would have been performed on the input sequence

prior to performing a temporal interpolation. Motion vector information could be used in such a process to obtain a better quality picture than would be obtainable with conventional techniques such as vertical-temporal filtering. However, the development of such a technique is a lengthy investigation on its own. Thus for the purpose of these experiments, the signal was assumed to be sequential, with account being taken of the half line vertical shift between fields. This amounted to assuming that the input signal contained no vertical frequencies above 156 cycles per picture height, so the output pictures produced were slightly soft vertically.

4.2.1 Details of the method used

All these investigations were performed using the 'basic' phase correlation method (as described in Section 3), with simple square windows on the input picture portions. Sub-pixel interpolation was performed on the correlation surface by first interpolating values for points for half integral pixel sites and then fitting a quadratic curve, as described earlier.

The input picture was divided up into non-overlapping blocks, 64 pixels horizontally by 32 lines vertically. A phase correlation surface was calculated between each block and the corresponding block in the field one picture period ago. The location of the three highest peaks in each correlation surface was calculated.

Vectors were assigned separately to each pixel; the menu of trial vectors for a given pixel consisted of the vectors measured in the block containing the pixel, as well as those measured in the immediately adjacent blocks. Thus for a pixel in a block in the middle of the picture, a maximum of 27 vectors would be tried. The reasoning behind the idea of using vectors measured in neighbouring blocks as well as the vectors measured in the block containing the pixel in question is as follows; if a small part of a moving object entered a given measuring block, its velocity may not be accurately measured in that block, whereas it would be accurately measured in the adjacent block containing the bulk of the object. Other situations where the 'sharing' of vectors would be advantageous include zoom movement, where the motion vector changes continuously across the picture. By including vectors from adjacent measuring blocks in the list of trial vectors, the assigned vector is not forced to change abruptly at the measuring block boundary.

The number of vectors in the trial vector menu that this technique gave was probably unnecessarily large. Refinements of the technique would probably limit this number by only using vectors corresponding to peaks in the correlation surface more than a given

threshold distance above the noise floor. The number of trial vectors could also be limited by only trying vectors that were different from each other by more than a certain threshold amount; thus when the velocity of an object was measured in two adjacent blocks and a nearly identical result obtained, only one of the two vectors would be tried.

In order to assign a motion vector to a pixel, an 'error surface' was calculated for each vector, as described in Section 3. A very simple algorithm was used to interpolate the input pictures so that non-integer vector lengths could be dealt with. This algorithm involved taking a weighted sum of the values of the four nearest pixels. This amounted to performing a linear interpolation horizontally followed by a similar interpolation vertically.

The error surface for each trial vector was filtered with a simple spatial filter with an aperture of the form

$$A(x, y) = \frac{1}{dx + dy + 1} \qquad (dx, dy \leqslant 2)$$

$$A(x, y) = 0 \qquad (dx, dy > 2)$$

where dx and dy are the absolute horizontal and vertical distances (in pixels) from the point (x,y) in the error array. This type of filter was used largely because it was easy to implement, and could probably be optimized further. A different form of filtering, such as median filtering, may prove to give better results.

A motion vector was assigned to every pixel; there was no upper limit set for the acceptable error. In an ideal implementation, pixels for which all vectors gave a large error would be investigated further, as discussed previously.

The luminance value of each pixel in the output picture was calculated by averaging the values in the adjacent two fields, at locations displaced by the motion vector for the pixel. Fig. 16 illustrates this idea. The simple two-dimensional linear interpolator described above was used to perform sub-pixel interpolation on the input fields. Although the frequency response of this interpolator was far from ideal, it was adequate to show if vectors were being assigned correctly.

4.2.2 Results of vector assignment investigations

Using the method outlined above, the even fields from the 'Vintage Car' test sequence described in Section 4.1.4. were generated from the odd fields.
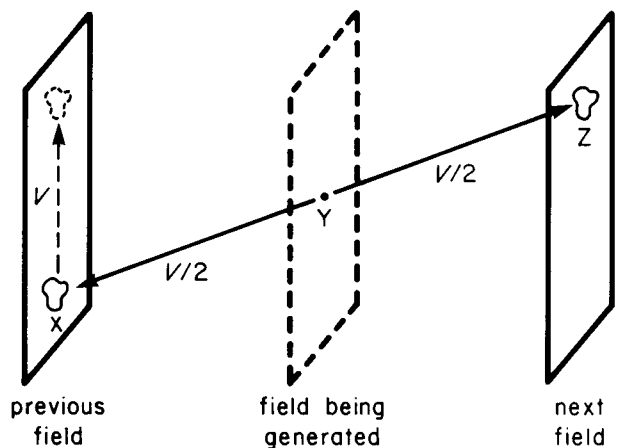

Fig. 16 - Generating an intermediate picture using simple motion compensated temporal interpolation.

Initial investigations performed without any spatial filtering on the error surface showed that incorrect vectors were often assigned to pixels due to noise (the test sequence is quite noisy). The use of the spatial filter discussed above cured this problem. The penalty for using a spatial filter of this kind is that the background immediately surrounding an object occasionally gets 'pulled along' with the object. It is likely that a median filter would give an improved performance.

When the error surface was spatially filtered, the interpolated pictures looked surprisingly good, with only one main problem remaining. Most parts of the interpolated picture appeared to be correct, except for the occasional disappearance of the silver surround at the bottom of the car's radiator and sections of the moving gate posts.

This problem was found to be due to large motion vectors being assigned to slowly moving areas in cases where both large and small vectors would be equally valid. Fig. 17 illustrates this problem in the case of the car's radiator. The stationary object is the
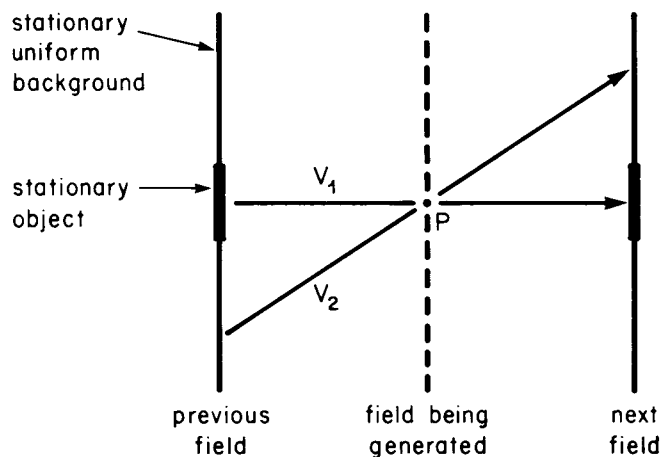

Fig. 17 - How a large and a small vector can both fit one point.

silver surround, and the uniform background is the dark area above and below it. The larger (erroneous) vector did not correspond to any movement present in the scene; it was due to a noise peak in one correlation surface. One possible cure for this problem would be to set a lower limit to the heights of peaks that were interpreted as real motion, as mentioned in Section 4.2.1. In the case of the disappearing gate post, the cause of the erroneous large vector was found to be more subtle. The gate was a periodic structure which moved about 4.7 pixels horizontally in a picture period, and had a horizontal repeat period of about 14 pixels (the spacing between centres of successive posts). This meant that there were two valid motion vectors for the gate, namely $+ 4.7$ and $- 9.3$ pixels per picture period (disregarding the effect of the edge of the gate). This situation is analogous to the well known 'reversing wagon wheels' effect, and is due to the temporal aliasing present in the signal. If the incorrect (larger) motion vector was chosen, the gate 'broke up' in the interpolated picture.

The problem was alleviated by multiplying the error surface for each vector by a function that increased with increasing vector length. This meant that when two vectors gave roughly the same match error, the shorter of the two vectors would be assigned. Several different forms of weighting function were tried; all worked reasonably well on the 'Vintage Car' test sequence. A wider range of test material would be required to find the exact form which gave the best results generally.

Once this modification had been incorporated, very presentable output pictures were generated.

Fig. 18 shows one interpolated field from the test sequence, and compares it to linear interpolation without the use of motion compensation. The improvement gained by the use of motion compensation is clearly visible, particularly on the moving gate.

As stated above, the vector measurement algorithm used in these investigations divided the input picture into 16 measurement blocks, measured up to three vectors per block, and 'borrowed' vectors from neighbouring blocks when forming a list of trial vectors for a given pixel. In order to see the effects of using fewer larger measuring blocks, an experiment was tried whereby the whole of the (quarter size) input picture was treated as one block, and four vectors were extracted from the resulting correlation surface. Somewhat surprisingly, the resulting interpolated pictures were almost indistinguishable from those obtained using many measuring blocks. However, the test sequence used was rather special, in so far as it only showed three principal moving objects (the gate, the car and the background), which moved largely as rigid bodies. A more general sequence containing many fast moving objects, rotations or zooms would probably look much better if more measuring blocks were used. Unfortunately, such test sequences were not available at the time.

4.2.3 Possible improvements to the temporal interpolation algorithm

As it stands, this temporal interpolation process has shown that vectors were being measured and assigned accurately. If the algorithm were to be used to perform temporal interpolation in a real application

(a)   *(b)*

Fig. 18 - Interpolation of an image from preceding and following television fields
(a) without and (b) with motion compensation.

such as standards conversion, it might be worth incorporating some improvements.

The vertical resolution available in the output pictures could be improved by performing an interlaced-to-sequential conversion prior to carrying out the interpolation, as mentioned previously. Use of motion information in the conversion process should enable full vertical resolution to be maintained in areas moving horizontally as well as in stationary areas. Some loss of vertical resolution in vertically moving areas is inevitable even with motion compensation, since the picture is no longer being properly sampled[3].

The horizontal resolution could be improved by using a more sophisticated interpolator than the linear interpolator described previously. Some investigations using an interpolator based on fitting a cubic spline[15] showed that such a technique increased the resolution of the output pictures at the expense of a significant increase in the processing time. Interpolators with even more taps would probably be used in broadcast quality applications.

A more fundamental improvement could be made by introducing an algorithm to deal with uncovered and obscured background. Areas of the picture being interpolated that contain uncovered background can be detected in several ways. Firstly, it is likely that a motion vector that gives a low match error value will not be found for such areas. Secondly, from 'optic flow' considerations, there must be some obscured or uncovered background at the boundary between regions with different vector components normal to the boundary.

Regions near a boundary where there is a vector component pointing away from the boundary will contain uncovered background. The picture

information that should be placed in this area in the interpolated picture will only be found in pictures taken at times after the time corresponding to the picture being interpolated. The area of the following picture that contains the required information can be found by examining the next but one picture, and following motion vectors back to find out which area of the picture originated from the area in question. Fig. 19 illustrates this process. Obscured background could be dealt with in the same way, but using preceding rather than following pictures.

4.2.4 Summary of motion vector assignment investigations

The experiments described above showed that it was possible to correctly assign motion vectors on a pixel-by-pixel basis using the method described in Section 3. The only modification found to be necessary was the incorporation of a weighting factor to bias the vector assignment towards small vectors. The technique could easily be extended to assign vectors on a block-by-block basis if required.

The performance of the technique as a whole appeared to give results which were probably significantly better than could be obtained with conventional vector measurement techniques. This is largely because of the pixel-by-pixel nature of the vector assignment, which allowed the boundaries of regions of the picture that were dealt with in different ways to closely follow object boundaries rather than block boundaries. It is also interesting to note that when large measuring blocks are used, the vectors measured by this technique are inherently limited to those corresponding to the major objects in the scene. Far from being a limitation, this has shown to be a desirable feature in many cases[9].
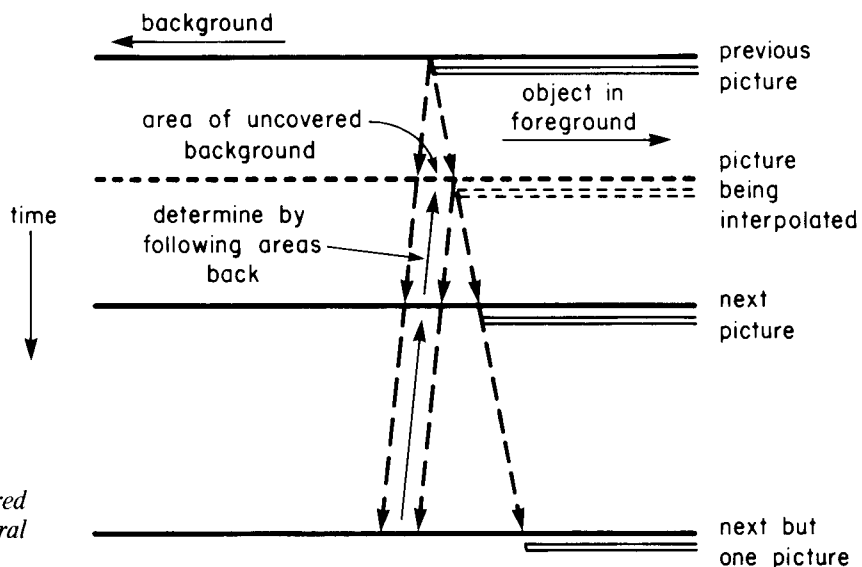

Fig. 19 - A method for dealing with uncovered background in a motion compensated temporal interpolation process.

Various improvements have been suggested over the basic technique outlined in Section 3. In order to determine which of these might be worth incorporating, it would be necessary to perform further simulation work on a larger range of test sequences. The exact form of the technique used would also depend on the use to which the vector information was to be put.

4.3 Application to interlaced picture sources

The simulation work described in Section 4.2 involved measuring motion vectors in an interlaced picture sequence. Since the aim of the investigation was to generate the odd interlaced fields of a sequence given only the even fields, the odd fields were not used in the vector measurement process. Thus all measurements occurred across a picture period rather than a field period. In a real application, it is likely that both interlaced fields would be available. However, further investigations have shown that it is still generally better to use only every other field in the vector measurement process. The reason for this is as follows.

If an odd field and an even field are used to measure motion vectors, any vertical frequencies above 156 lines per picture height (for a 625-line interlaced system) would contribute noise to the correlation surface, except in cases where there is vertical movement of nearly an odd number of picture lines per field period. The vertical aliasing produced by these frequencies would also hamper the vector assignment process.

If, however, two fields of the same type are used, then the vertical movement speeds which do not cause trouble are near even numbers of picture lines per field period. Hence stationary vertical detail does not confuse the measuring system. Vertical detail moving vertically still presents a problem, but as such detail has probably not been sampled correctly by the interlaced source it is often difficult to deal with it in subsequent processing. In other words, the measuring system can be optimized for zero vertical velocity, which is by far the most useful velocity to be able to deal with best.

This approach also avoids the problem of the spurious half-line vertical shift which is introduced when measurements are performed between odd and even fields.

Since this approach involves making measurements across a picture period, all velocities are doubled. This can be advantageous since it allows a more accurate measurement to be made. It also means that two similar velocities will be distinguished more

easily. However, it may be necessary to be able to deal with higher velocity components, maybe corresponding to shifts of 30 pixels or more.

If motion vectors are required for the intervening fields, these can be interpolated from the vectors measured on either side. In order to perform this interpolation, it is necessary to assume that objects do not accelerate appreciably in one picture period. This assumption is nearly always justified.

5. INVESTIGATION OF APPLICATIONS FOR MOTION VECTOR MEASUREMENT TECHNIQUES

Motion vector measurement techniques have many applications in the field of television broadcasting. Several applications of the extended phase correlation technique have been investigated by simulation, and are reviewed below.

5.1 The use of motion vector information in a bandwidth reduction system using DATV

Knowledge of the motion present in a television picture can be a great help in many bandwidth reduction processes since it can provide the key to removing much of the redundancy in the signal. The differences between successive television pictures are often largely due to movement, so it is generally possible to reconstruct most of a picture given only the preceding picture and information about the motion content.

One particular form of bandwidth reduction system that can benefit from motion vector information is described in Refs. 1, 2, 3. This bandwidth reduction system is based on subsampling the input signal in order to provide a 4:1 bandwidth reduction. Two different types of pre-filter can be applied to the signal prior to subsampling; one optimum for stationary areas, the other optimum for moving ones. In stationary areas, the decoder uses four fields of samples to reconstruct an image with high spatial resolution but poor temporal resolution. In moving areas, one field of samples is used to reconstruct each image, resulting in reduced spatial resolution but good temporal resolution.

The reduced bandwidth signal produced by this system consists of an analogue part containing the values of the subsamples, and a digital part that tells the decoder which type of reconstruction algorithm to use for each part of the picture. For this purpose the picture is divided into small blocks, each block being sent in one of the two modes. A television transmission system such as this, which uses both analogue picture data and digital 'assistance' data, has been termed Digitally Assisted Television, or DATV.

Although this bandwidth reduction system was found to work well from the point of view of artefacts, the loss of resolution in moving areas which the observer's eye could track was objectionable. In theory it is possible to extend the high spatial resolution mode into areas of the picture containing movement, if the corresponding motion vectors can be measured. The prefilter apertures can be distorted and the subsampling lattice moved in such a way as to render the image stationary as far as the bandwidth reduction system is concerned. Motion vectors can be sent to the decoder using the digital part of the signal, to enable the signal to be reconstructed with moving objects shown in their correct positions and at a high spatial resolution.

Simulations have been performed to determine the effectiveness of motion compensation in improving the performance of this bandwidth reduction system[3]. The extended phase correlation technique was used to assign a motion vector to each small block in the picture. This technique is particularly suited to this application for several reasons. One reason is that it is necessary to measure motion vectors to sub-pixel accuracy in order to minimize the 'alignment error' between the four fields used to reconstruct the image. As shown earlier, this vector measurement method is capable of accuracies of the order of 0.1 pixel, which is easily good enough for this application. Another reason is that in order to keep the bandwidth of the digital assistance channel as low as possible, the number of different vectors that are assigned needs to be limited. The extended phase correlation technique inherently produces a limited menu of trial vectors for any given picture area. Thus the vector information can be sent to the decoder in the form of this menu, with just a few bits per block to indicate which particular vector was assigned.

The simulation work has shown that the performance of the bandwidth reduction system can be significantly improved by the addition of motion compensation. Further details of results are given in Ref. 3.

5.2 Applications based on temporal interpolation

There are several applications for motion vector measurement techniques that require intermediate pictures (or fields) to be generated in a moving sequence. Processes such as standards conversion between 50 Hz and 60 Hz field rates, which have previously been performed using linear temporal interpolation[16], can benefit significantly from the use of motion compensation. More demanding temporal standards conversion operations (such as the generation of intermediate fields for the display of film or slow motion sequences, or for display field rate up-

conversion) cannot be performed satisfactorily using linear temporal interpolation, since moving objects become unacceptably blurred. In these cases, additional fields are usually obtained by simply repeating fields as many times as required. The use of motion compensated temporal interpolation can enable intermediate fields to be generated with moving objects correctly positioned and without blurring.

The simulation program used to obtain the results presented in Section 4 was extended to enable any number of intermediate fields to be generated between each pair of input fields. This enabled all the applications discussed above to be investigated.

The program was modified to perform a motion compensated interlaced-to-sequential conversion on each input field prior to performing the temporal interpolation. The conversion was achieved by interpolating the missing lines in each field from the luminance levels in the preceding and following fields at locations displaced by the appropriate motion vector. This required spatial interpolation to be performed on the adjacent fields. This method allows full vertical resolution to be retained in areas with no vertical movement (and areas moving at a vertical speed of an even number of picture lines per field period). The vertical resolution available drops to half its theoretical maximum in areas moving at a vertical speed of an odd number of picture lines per field period; this is unavoidable due to the interlaced nature of the source. Horizontal motion does not affect the spatial resolution available. This process is capable of generating a high quality sequential picture.

A number of different temporal standards conversion simulations were performed, using the motion vector measurement algorithm described in Section 4 to assign vectors on a pixel-by-pixel basis. These included 50 Hz to 60 Hz standards conversion, up-conversion from 50 Hz to 75 Hz, generation of slow motion sequences at 20% of normal speed, and improved display of film motion. Despite the lack of an algorithm to deal with uncovered background as discussed in Section 4.2.3, the output pictures showed minimal impairments. The quality of the generated sequence was largely independent of the ratio of the number of input pictures to output pictures. Simulation work using a wider range of test sequences is required to fully evaluate this technique, although results with the limited test material available were encouraging.

6. HARDWARE IMPLEMENTATION OF THE EXTENDED PHASE CORRELATION TECHNIQUE

This Section outlines a possible way of implementing the extended phase correlation technique described above.

Fig. 20 shows a block diagram of a possible hardware implementation. Two-dimensional Fourier transforms are performed on the incoming digitized video signal, on a block-by-block basis. The size of the blocks would depend on the application, but typically a block might be 64 pixels by 64 lines, as used in the simulation work. Only one field of samples would probably be used, for the reasons discussed in Section 4.3. This arrangement would require a total of about 44 blocks, requiring maybe 5 circuits performing Fast Fourier Transforms (FFTs), assuming it takes 2 clock cycles to perform each 'butterfly' operation within the FFT.

The transformed data is passed to a circuit which computes the phase difference between each frequency component and the corresponding component from the previous picture. This circuit could include a simple filter as discussed in Section 4.1.3.

The phase difference information is passed to another set of FFT circuits, identical to those used previously. Some saving in hardware could be achieved by using the fact that the output from these transforms consists only of real numbers, since it is possible to perform two such transforms as one complex transform by using the symmetry properties of Fourier transforms on real data. A similar saving could be made with the first set of transform circuits. A further saving in hardware could also be achieved by quantization of the phase angles prior to transformation[10]. In practice however, it may be better to perform all transforms as complex manipulations to maybe 12 or 16 bit accuracy, due to the availability of relatively cheap VLSI multipliers. This would allow the design itself to be kept simple.

The output from the second set of FFT circuits consists of a number of correlation surfaces, one for
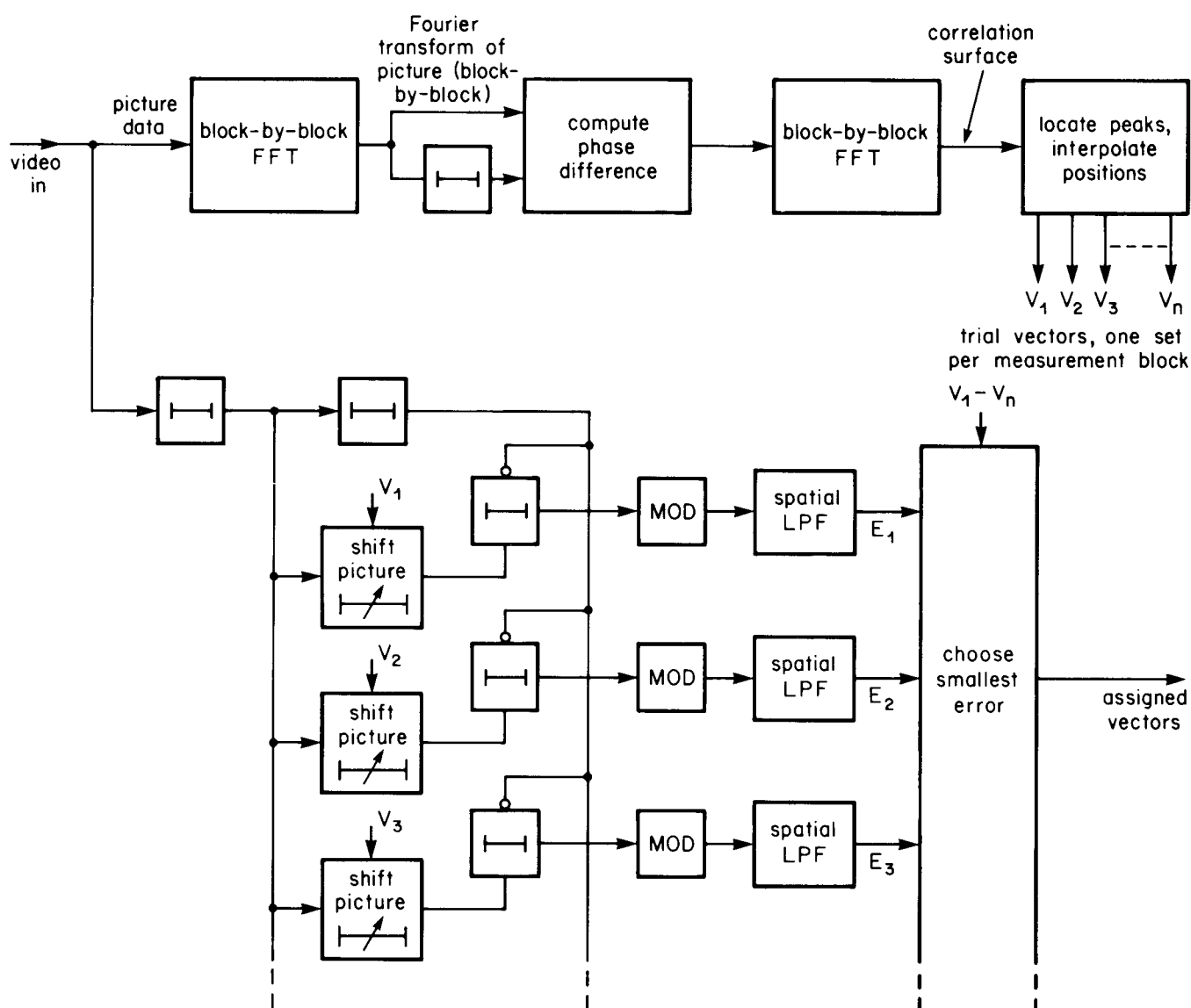

Fig. 20 - The block digram of a possible hardware implementation of the extended phase correlation technique for motion vector measurement.

each measurement block. These surfaces are interrogated by a number of fast microprocessors, which determine the locations of the dominant peaks and perform sub-pixel interpolation to find their exact location. These processors produce a list of trial vectors for each measuring block.

Each trial vector is sent to a variable shift circuit, which displaces the input picture by the given vector. Such a circuit is implemented as a variable delay, with the addition of simple sub-pixel interpolation to allow non-integer vectors to be dealt with correctly. A set of subtractors computes the difference between each shifted picture and the previous input picture, producing an 'error surface' as discussed in Section 3.

The magnitude of this surface is filtered with a low pass filter to reduce the effect of noise, and multiplied by a scaling factor dependent on vector length, as described in Section 4.2.2. The filtered and scaled error signals are passed to a circuit which selects the smallest error for each area of the picture. The vector corresponding to this error becomes the final output from the equipment.

If vector measurement was being performed between two fields one picture period apart, it would be possible to make one shift circuit test two vectors, one during each field period. The number of vectors to be tried would depend on the application, but a total of 12 vectors (requiring 6 shift circuits) would probably give a good performance even in critical applications. It is estimated that the total circuitry of the vector measurement equipment would occupy three standard '19 inch' racks approximately 270 mm high.

7. CONCLUSIONS

In an effort to find a highly effective technique to measure motion in television pictures, many published papers have been studied. None appear to report a technique totally suitable for critical applications such as broadcast quality temporal interpolation.

An extension to one basic method (phase correlation) has been devised, which appears to overcome many of the problems of more conventional techniques. This technique has been investigated by computer simulation and found to be most effective.

There are numerous applications for such a vector measurement technique. Applications including bandwidth reduction, standards conversion, the improvement of slow motion portrayal and film motion portrayal have been investigated with encouraging results.

One possible way of implementing the extended phase correlation technique in hardware has been outlined. The amount of hardware required would be of the order of three '19 inch' racks, each about 270mm high.

Thus a motion vector measurement technique suitable for many critical applications has been devised and tested by simulation. In terms of accuracy and size of measurable movements, it appears to perform better than other published techniques. Hardware to implement the technique in real time is also practicable and will be constructed to confirm the effectiveness of this technique on a wide range of subject matter.

8. REFERENCES

1. STOREY, R. 1986. HDTV Motion Adaptive Bandwidth Reduction using DATV. BBC Research Department Report No. BBC RD 1986/5

2. STOREY, R. 1986. HDTV Motion Adaptive Bandwidth Reduction Systems and Their Interplay with Source and Display Scanning Structures. 11th International Broadcasting Convention IBC 86. Brighton. 19th-23rd September 1986. IEE Conference Publication No. 268 pp 167-172.

3. THOMAS, G.A. 1986. HDTV Bandwidth Reduction by Adaptive Subsampling and Motion Compensation DATV Techniques. SMPTE Journal, Vol. 96, No. 5, pt. 1, May 1987, pp 460-465.

4. NETRAVALI, A.N., ROBBINS, J.D. 1978. Motion Compensated Television Coding: Part 1. The Bell System Technical Journal Vol. 58, No. 3, March 1979, pp 631-670.

5. SABRI, S. 1983. Movement Compensated Interframe Prediction for NTSC Colour TV Signals. IEEE Trans. Commun. Vol. COM-32, No. 8, August 1984, pp 954-968.

6. WELLS, S.C. 1984. Motion Estimation Algorithms and their Application to Hybrid Coding. IEE Colloquium on Motion Adaptive Video Processing, May 1984. IEE Digest No. 1984/59.

7. NINOMIYA, Y., OHTSUKA, Y. 1981. A Motion-Compensated Interframe Coding Scheme for Television Pictures. IEEE Trans. Commun. Vol. COM-30, No. 1, January 1982, pp 201-211.

8. JAIN, J.R., JAIN, A.K. 1981. Displacement Measurement and its Application in Interframe Image Coding. IEEE Trans. Commun. Vol. COM-29, No. 12, December 1981, pp 1799-1808.

9. FURUKAWA, A., KOGA, T., LINUMA, K. 1984. Motion-Adaptive Interpolation for Videoconference Pictures. Links for the Future: Science, Systems and Services for Communicaions, Ed. Dewilde and May, IEEE International Conference on Communications, May 14-17. IEEE/Elsevier Science Publishers B.V., 1984, Amsterdam, pp 707-710.

10. PEARSON, J.J., HINES, D.C., GOLOSMAN, S., KUGLIN, C.D. 1977. Video-rate Image Correlation Processor. S.P.I.E. Vol. 119. Application of Digital Image Processing (IOCC 1977), pp 197-205.

11. TIAN, Q., HUHNS, M.N. 1985. Algorithms for Subpixel Registration. Computer Vis., Graph, Im. Proc. Vol. 35, No. 2, 1986, pp 220-233.

12. BUXTON, B.F., BUXTON, H., MURRAY, D.W., WILLIAMS, N.S. 1985. Machine Perception of Visual Motion. GEC J. Res., Vol. 3 No. 3, 1985, pp 145-161.

13. SMITH L.D.R., ROBINSON, G.P., STEVENSON, D.J., SANGHERA, H., QUARENDON, P. 1986. A Semi-Automatic Computerised System for the Production of Quantitative Moving Three Dimensional Images of the Heart and Coronary Arteries. IEE Second International Conference on Image Processing and its Applications, June 24-26, 1986. IEE Conference Publication No. 265, pp 72-76.

14. D'HAEYER, J., BRUYLAND, I. 1986. A New Fast Motion Estimation Algorithm and its Application in Image Coding. EURASIP International Workshop on the Coding of HDTV. November 1986. European Association for Signal Processing publication Coding for HDTV, Vol.1.

15. HOU, S.C., ANDREWS, H.C. 1978. Cubic Splines for Image Interpolation and Digital Filtering. IEEE Trans. ASSP Vol. ASSP-26, No. 6, December 1978, pp 508-517.

16. CLARKE, C. K. P., TANTON, N. E. 1984. Digital Standards Conversion: Interpolation Theory and Aperture Synthesis. BBC Research Department Report No. BBC RD 1984/20.